ABSTRACT
This paper summarizes the results of the CNS Summit Data Quality Monitoring Workgroup analysis of current data quality monitoring techniques used in central nervous system (CNS) clinical trials. Based on audience polls conducted at the CNS Summit 2014, the panel determined that current techniques used to monitor data and quality in clinical trials are broad, uncontrolled, and lack independent verification. The majority of those polled endorse the value of monitoring data. Case examples of current data quality methodology are presented and discussed. Perspectives of pharmaceutical companies and trial sites regarding data quality monitoring are presented. Potential future developments in CNS data quality monitoring are described. Increased utilization of biomarkers as objective outcomes and for patient selection is considered to be the most impactful development in data quality monitoring over the next 10 years. Additional future outcome measures and patient selection approaches are discussed.

INTRODUCTION
There is a dearth of empirically based, peer-reviewed literature critiquing methodologies for monitoring and enhancing data quality in central nervous system (CNS) clinical trials. Moreover, there is little published guidance for
Matching these products to specific clinical trial designs and subject populations. To address these issues, we convened a panel at the 2014 CNS Summit in Boca Raton, Florida, that included prominent representatives of the pharmaceutical industry, clinical trials sites, and companies that provide specialized services targeting enhancement of outcomes using subjective CNS efficacy measures. The panel was charged with critiquing existing data quality techniques and recommending future directions for enhancing data quality.

HAS DATA QUALITY MONITORING IN CLINICAL TRIALS BEEN WORTH IT?

At relevant points during the 2014 CNS Summit, the audience, which comprised 350 representatives of the same constituencies as the panel, was polled utilizing an audience response system (ARS). When asked, “What is the current level of satisfaction with data quality monitoring in CNS trials?” the audience was nearly evenly split between “satisfied” and “dissatisfied” (Figure 1).

As shown by the examples in Figure 2, the range of data monitoring and quality assurance techniques is quite broad. However, the evidence in support of one intervention over the other is sparse, often uncontrolled, and lacking in independent verification. Moreover, the potential negative effects of these techniques have not been systematically explored. Examples of potentially untoward effects include skewing patient selection, loss of systematic variability with regression toward the mean, and increased site burden.

The polling results in Figure 3 demonstrate that a large majority of those polled (94%) endorsed the value of data quality monitoring.

CURRENT DATA QUALITY MONITORING TECHNIQUES

To provide a basis for the panel’s critique of data quality methodologies, each member of the data quality service companies
represented on the panel provided two anonymized case examples of data quality methodology. The combined set could be divided into three broad categories: 1) diagnostic confirmation and study eligibility; 2) data monitoring and ratings accuracy; and 3) potential adverse effects of surveillance. The following slides were presented and discussed by Dr. David Walling:

1. **Diagnostic confirmation and study eligibility.** Figure 4 illustrates a process of external review of site rater-completed forms documenting that subjects met eligibility criteria at screening. The 14-percent screen fail rate based upon the eligibility review at the beginning of the study fell to approximately six percent, consistent with a learning effect and resulting in submission of more appropriate subjects.

   Figure 5 illustrates an approach in which subjects complete a computerized assessment (bipolarity index [BPI]) of five domains germane to diagnostic validity and suitability as clinical trial subjects. In the current example, concerns about family histories and responses to prior treatment were not resolved by follow-up calls to the sites.

2. **Data quality monitoring.**

   Figure 6 illustrates the relationship between the length (7–59 minutes) of recorded Brief Psychiatric Rating Scale (BPRS) interviews (n=392) and variance between the ratings of the site investigators (n=43) and those of independent external reviewers.

   The 21-minute point was associated with the greatest concordance between the site and independent raters and very long and very short interviews with the lowest concordance.

   Figure 7 describes a methodology characterized by an unobtrusive centralized expert review of blinded data for patterns associated with poor quality. Expert feedback was provided to the site as required. The number of required site contacts decreased over the first four months,
and the improvement persisted through 14 months.

Figure 8 describes a process of evaluation of the case report form (CRF) interview and scoring results based on independent review of audio or video recordings of the assessment. Errors were detected in 30 percent and 12 percent of the assessments by audio and video review, respectively. Notably, some errors, such as leaving the word list visible to the subject, could only be detected by video review.

Figure 9 illustrates that multiple contacts between sites and clinical resource organizations are required for agreed upon error corrections to be entered into the electronic data capture (EDC). In this sample, two and three contacts were required to reduce the level of outstanding errors to 40 percent and four percent, respectively.

Figure 10 illustrates an association between data quality monitoring and remediation and a reduction in the incidence of data quality problems in three different therapeutic indications. In the major depression study, the incidence of problems such as identical scoring across visits and inconsistencies between the Montgomery–Åsberg Depression Rating Scale (MADRS) and Hamilton Rating Scale for Depression (HAM-D) fell from 60 percent to 38 percent. In the Parkinson’s disease example, the incidence of errors in on/off diaries fell from approximately two percent to less than one percent (0.8%). In the Alzheimer’s disease study, the proportion of visits with at least one error fell from 94 percent at the initiation of the study to 69 percent approximately 18 months later.

Collectively, the three examples suggest that sites enhance the quality of their rating behaviors in response to feedback from data quality monitoring.

Figure 11 illustrates comparison of observed and expected correlations among scales as a quality indicator in an Alzheimer’s disease clinical trial.

Figure 12 illustrates a dashboard in
which sites are ranked with respect to their risk to study quality. The components and weighting of quality factors comprising the composite index is determined by consensus between the sponsor and vendor.

In contrast to the study level dashboard in Figure 12, Figure 13 illustrates the performance of a single site relative to other sites in the study on a range of quality factors. The site shown was an outlier on the proportion of consecutive visits scored identically on the Unified Parkinson’s Disease Rating Scale (UPDRS) and on a measure of clustering at baseline around inclusion scores.

Figure 14 illustrates a retrospective analysis of a Phase III depression study in which separation between study drug and placebo became evident when sites with significant quality issues were removed. This data mining exercise illustrates the importance of data quality in signal detection.

Figure 15 illustrates the results of asking eastern European subjects at the end of a psychometric study if they would have participated in the interview had it been recorded. In the outpatient sample, prominent negative symptoms were associated with negative views toward being videotaped.

In the inpatient sample, greater insight that one had schizophrenia (scoring lower on Item G12 of the Positive and Negative Syndrome Scale [PANSS]) was associated with negative attitudes toward videotaping.

PHARMACEUTICAL COMPANY PERSPECTIVE ON DATA QUALITY MONITORING

The consensus among sponsor representatives was that data quality monitoring would be a permanent and growing part of the clinical trials landscape and would have large positive effects on signal detection. However, these panel members noted the absence of an empirically based methodology for matching the type and quantity of data quality measures to specific study designs to maximize return on investment. While helpful,
monitoring techniques require continuous innovation and refinement to address new study designs, continuing quality issues, and delays in entry of corrections. The value of partnership among data quality monitors, sites, and patients was seen as essential to successful trial execution. It was noted that special adaptations are needed for effectiveness trials, which may have less frequent data acquisition and greater reliance on patient-reported outcomes (PROs) than explanatory trials.

**SITE PERSPECTIVE ON DATA QUALITY MONITORING**

The site representatives affirmed the view that data quality monitoring and remediation benefit signal detection. These panel members noted that although the second set of eyes of the independent reviewer is welcomed, the process should transpire in a collegial, collaborative manner. More and faster feedback to the site on interview and ratings quality would be appreciated. Research to identify the most effective and efficient data quality monitoring techniques is needed. Data monitoring should be operationalized to minimize disruption of the flow of procedures at the site. Many time- and labor-consuming data monitoring procedures are uncompensated.

**FUTURE DIRECTIONS IN DATA QUALITY MONITORING**

Polling of the meeting audience revealed that the most impactful developments in CNS data quality monitoring in the next decade were expected to arise from increased utilization of biomarkers as objective outcomes (Figure 16).

Representatives of the data quality service companies added that genotypic and phenotypic patient characteristics would increasingly be utilized as clinical trial subject selection tools and outcome measures, but would require validation against relevant behavioral measures and functional outcome. Increasingly sophisticated analytic techniques for large data bases would assist this process.

A number of future outcome measures and approaches were cited as likely to become more prominent, including biomarkers; computerized, adaptive assessments; shorter, more...
focused subjective scales; domain- and dimension-based measurement outcomes of greater relevance to patients and their caregivers; physiological-, physical-, and home-based measures; and more relevant and culturally sensitive functional assessments. There was no agreement on how to prioritize development or introduction of these methodologies.

Creation of data banks of rater experience, credentials, and performance was agreed upon as high priority, achievable pre-competitive objectives.

There was broad consensus for the need to systematically and empirically assess current data quality techniques to determine their value, and to inform the type and amount of intervention appropriate for a given clinical trial design.

**CONCLUSION**

In a pre-competitive session among pharmaceutical companies, clinical service companies, and clinical trial sites, there was a strong consensus that data quality monitoring and remediation has large positive effects on signal detection in CNS trials. A sense of partnership and commonality of mission prevailed among the constituencies. Data supporting the ability of a broad range of methodologies to detect and remediate errors were presented. However, a dearth of controlled comparisons, rarity of independent verification, and lack of publication of negative results limited assessment of their value. It is hoped that in the future, empirically based methodologies will be developed to match the type and quantity of data quality measures to specific study designs. Such methodologies are needed to assess the cost and benefit of monitoring and interventions. There was wide consensus that, in the future, more objectively measured endpoints would be validated to replace many of the current subjective endpoints in CNS clinical trials.